ARL-TR-8301 ● Feb 2018

US Army Research Laboratory

Human Factors Evaluation of Advanced Video Activity Analytics (AVAA) Functionality

by Kristin M Schweitzer, Anthony J Ries, Patricia L McDermott, Beth M Plott, Elizabeth A Wilson, and Greg P Morrow

NOTICES

Disclaimers

ARL-TR-8301 ● FEB 2018

US Army Research Laboratory

Human Factors Evaluation of Advanced Video Activity Analytics (AVAA) Functionality

by Kristin M Schweitzer and Anthony J Ries
Human Research and Engineering Directorate, ARL

Patricia L McDermott and Elizabeth A Wilson
MITRE, McLean, VA

Beth M Plott
Alion Science and Technology, McLean, VA

Greg P Morrow
CACI, Arlington, VA

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
February 2018	Technical Report	6 January 2016–11 January 2017

4. TITLE AND SUBTITLE Human Factors Evaluation of Advanced Video Activity Analytics (AVAA) Functionality	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Kristin M Schweitzer, Anthony J Ries, Patricia L McDermott, Beth M Plott, Elizabeth A Wilson, and Greg P Morrow	**5d. PROJECT NUMBER** 0001
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory Human Research and Engineering Directorate (ATTN: RDRL-HRB-DE) ARL South at San Antonio North Paseo Building, Room 2.212 One UTSA Circle San Antonio, TX 78249	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-8301
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The Advanced Video Activity Analytics (AVAA) system is a full-motion video exploitation capability designed to reduce imagery analyst workload and enable faster and more-accurate production of intelligence products. Phase II of the human factors evaluation examined the impact an automated person–vehicle–object (PVO) computer vision algorithm pipeline had on imagery analysts’ cognitive workload and performance during realistic scenario-based operations. The results from multiple measures indicate that AVAA’s PVO pipeline did not significantly affect the measures analyzed for performance, cognitive workload, or situation awareness as compared with the baseline AVAA system. However, analysts identified multiple usability recommendations for AVAA that would improve usability and follow-on development.

15. SUBJECT TERMS

full-motion video, FMV, imagery, exploitation, analyst, intelligence, human factors, engineering, HFE, computer vision analytic, CVA, annotation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Kristin M Schweitzer
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified	UU	50	(210) 458-6839

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Summary

The Advanced Video Activity Analytics (AVAA) system is a full-motion video (FMV) exploitation capability. AVAA's objective is to reduce an analyst's workload and enable faster and more-accurate production of intelligence products.[1] To evaluate progress in the development of AVAA's capability to reduce analytical burden on FMV analysts, we conducted a series of human factors evaluations of the software's functionality. The purpose was to ensure development was guided with repeated user feedback and to establish baseline performance so we could then determine whether AVAA functionality reduced analyst workload.

Technical descriptions of the scalable architecture and computer vision (CV) algorithm pipelines used to provide the functionality evaluated in this study have been published previously.[2,3] Generally speaking, the user events utilized 4 CV pipelines. The Phase I evaluation used Data Extraction (for key-length-value [KLV]) decoding), KLV Demux and Improvement (for sensor connection and KLV editing), and Video National Imagery Interpretation Rating Scale (V-NIIRS) Video Quality pipelines. The Phase II evaluation included an additional PVO (person–vehicle–object) Detect, Classify, Track CV pipeline.

In Phase I, AVAA developers achieved forensic processing, exploitation, and dissemination capability of electro-optical and IR FMV at scale. This included the system capability to automatically rate the image quality of the video using V-NIIRS and the ability to search for FMV with specific V-NIIRS ratings. We conducted an initial evaluation with experienced imagery analysts under Project No. ARL 14-020. Findings from the initial evaluation suggested that analysts were able to identify more targets with the V-NIIRS filter than the baseline condition in time-pressured situations.[4] The initial evaluation successfully demonstrated a multi-aspect approach to estimate operator functional state during system evaluation.

In Phase II, AVAA developers achieved the capability to automatically indicate possible areas or activities of interest on saved video. The analyst could search for

---

[1] Swett B. Advanced Video Activity Analytics (AVAA) overview. AVAA Preliminary Design Review presentations; 2013 Nov 6–7; Lorton, VA.

[2] Thissell WR, Czajkowski R, Schrenk F, Selway T, Ries AJ, Patel S, Palaniappan K. A scalable architecture for operational FMV exploitation. Proceedings of the International Conference on Computer Vision; 2015 Dec; Santiago, Chile. p. 10–18.

[3] Wilson E, Patel U. Cognitive load assessment for intelligence analysts through FMV analytics. Proceedings of the Interservice/Industry Training, Simulation, and Education Conference; 2015 Dec; Orlando, FL. p. 1–12.

[4] McDermott PL, Plott BM, Ries AJ, Touryan J, Barnes MJ, Schweitzer KM. Advanced Video Activity Analytics (AVAA): human factors evaluation. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015. Report No.: ARL-TR-7286.

annotated video and could either accept or reject the automated annotations as accurate, edit existing annotations, and create new ones.

This report describes the Phase II user event, which was designed to empirically evaluate the impact of AVAA on user performance and workload. The focus was to determine whether the CV pipeline for an automated PVO detector imparted an undue burden on the user during realistic scenario-based operations. We examined user actions, verbal feedback, and subjective measures for cognitive workload, situation awareness, and usability. We also recorded physiological measures such as electroencephalography data, electrocardiogram data, saccade magnitude, and blink duration.

A secondary evaluation effort in Phase II modeled AVAA's impact on the cognitive workload of imagery analysts with US Army military occupational specialty 35G.[5]

[5] Plott BM, McDermott PL, Barnes MJ. Advanced Video Activity Analytics (AVAA): human performance model report. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2017.

1. Introduction

Advanced Video Activity Analytics (AVAA) is designed to relieve the human factors burden on the analyst, reduce the time it takes him/her to analyze video (and thus increase the amount of video one analyst can exploit), and improve the ability of the operator to locate targets accurately within the videos while incorporating good usability principles.

To determine whether AVAA's current functionality was meeting these capability requirements, we designed a user event to capture usability and imagery analyst performance from multiple aspects. We developed 2 realistic scenarios for imagery analysts to complete and then recorded the number of button clicks they made for various actions and their total time on task during full-motion video (FMV) exploitation. We graded their responses on end-of-mission reports to examine their ability to derive applicable information from the videos. Users completed the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) to estimate subjective workload, the Situation Awareness Rating Technique (SART) to gauge perceived situational awareness, and a usability questionnaire to describe perceived functionality. Data collectors recorded their observations and each analyst's verbal feedback. For a subset of volunteers we evaluated measures such as electroencephalography (EEG) power ratios, accuracy and reaction time for a secondary auditory task, heart-rate variability, and eye tracking data to further quantify the burden AVAA imposes upon its users. The physiological measures and the rationale for their inclusion in this evaluation are detailed in the following.

Testers and evaluators have traditionally relied upon self-assessments to estimate the cognitive state of a user; however, they often break task flow by interrupting the user at various times throughout the task, leading to disruptions in mental concentration. Additionally, subjective self-assessments are not sensitive to fluctuations of cognitive state within a task; instead, they provide an average estimate over an interval of time, often on the order of several minutes, which can introduce unintended effects due to memory lapses (Moroney et al. 1995). To account for these shortcomings, self-assessments may be augmented with tools that measure continuous, objective metrics related to cognitive state and task performance. Multi-aspect measurements that draw upon multiple sources of information may provide the best approach for obtaining estimates of cognitive state.

Physiological and behavioral measurements obtained from EEG, electrocardiography (ECG), eye-tracking, and overt performance (e.g., reaction time and accuracy) have shown reliable, objective quantification of cognitive states primarily associated with workload and fatigue (Dinges and Powell 1985; Makeig and Inlow 1993; Dinges et al. 1998; Berka et al. 2007; Johnson et al. 2011; Stikic et al. 2011). EEG measures electrical output of the brain via electrodes attached to the scalp, where frontal theta power is typically linked with working memory processes, and posterior alpha power is typically linked to inhibitory processes. Increased workload is commonly associated with increases in frontal theta power (5–7 Hz) over the frontal midline and the posterior parietal electrodes and concomitant decreases in posterior alpha power (8–12 Hz) (Gundel and Wilson 1992; Brookings et al. 1996; Gevins et al. 1998; Fournier et al. 1999; Gevins and Smith 2003; Fairclough et al. 2005; Brouwer et al. 2012).

Other studies have shown that event-related potential (ERP) components ranging from early sensory responses like the auditory-evoked N100 (negative amplitude 100 ms after auditory event being measured) to components reflecting higher-order cognitive processes like the P300 (positive amplitude 300 ms after event) are sensitive to changes in workload. Typically, the ERPs are evoked using “probe” stimuli in a secondary task. The ERPs often elicit smaller evoked responses during high cognitive workload conditions with the primary task than low levels of cognitive workload with the primary task (Kramer et al. 1995; Allison and Polich 2008; Miller et al. 2011).

ECG measures such as heart rate and heart rate variability are also sensitive to changes in cognitive states related to workload. Typically, tasks that increase resource allocation from working memory demands show a concomitant increase in heart rate and decrease in heart rate variability (Mulder et al. 2004; Brookhuis and de Waard 2010).

Eye-tracking measurements also provide objective indices of user cognitive state. Research has shown that as task demands rise and cognitive workload increases, blink rate and blink duration decrease, and fixation frequency (number of fixations per time) increases (Van Orden et al. 2001; Wilson 2002; Ahlstrom and Friedman-Berg 2006). Others have observed changes in pupil diameter as a function of workload, noting increases in pupil diameter as workload increases (Backs and Walrath 1992; Van Orden et al. 2001).

Workload has been defined as the relationship between the processing capacity of the user and the processing requirements of the task (O’Donnell and Eggemeier 1986; Hart and Staveland 1988). It is limited by the available processing resources in the brain, coupled with task demands, and fluctuates based on the quantity of concurrent information processing in working memory. In the current evaluation, neural, ocular, cardiac, behavioral, and subjective data were used to estimate cognitive workload for the purpose of assessing cognitive demands induced by the AVAA capability set.

2. Methods, Assumptions, and Procedures

The evaluation used a counterbalanced, within-subjects design that compared AVAA functionality with automated PVO detections ("PVO" condition) to AVAA functionality without automated detections ("Baseline" condition). We provided 2 scenarios involving improvised explosive devices (described in the following). All analysts completed the Vest scenario first and the Vehicle scenario second. We counterbalanced the order of presentation of the experiment conditions.

2.1 Participants

We recruited 27 volunteer participants with Army military occupational specialty (MOS) 35G (imagery analyst) experience from the 111th Military Intelligence Brigade and the New Systems Training and Integration Directorate. Median age was 34 years (minimum = 24 years, maximum = 55 years). Duty positions were active duty E6 (n = 15) and E7 (n = 4) and civilian (n = 8). All participants had at least 2 years of experience as a 35G, with a mean time in MOS of 9.9 years (standard deviation [SD] = 7.7 years) and a maximum time of 33 years. All also had experience with imagery analysis outside the continental United States, and 13 had imagery analyst training beyond the standard 35G MOS training. Eight participants volunteered to have their physiological and behavioral data recorded.

2.2 Apparatus

The US Army Intelligence Center of Excellence's Experimentation and Analysis Element at Fort Huachuca, Arizona, hosted 5 operator workstations. All 5 workstations consisted of a laptop computer for processing, a stand-alone screen for all data visualization, and a standard keyboard and mouse.

In addition to the basic workstation setup, the 2 EEG workstations had headphones and a 17-inch touchscreen monitor connected to a laptop to generate the auditory stimuli and process user responses. EEG and ECG data were acquired at a sampling rate of 256 Hz from the B-Alert x24 Wireless Sensor Headset, with the single-trial ERP montage using the B-Alert software package (Advanced Brain Monitoring, Carlsbad, California). EEG signals were sent via Bluetooth to an external syncing unit, which connected to the data acquisition laptop through USB. In addition to the scalp electrodes for EEG data collection, 2 external input channels were placed just below the participant’s right clavicle and below the left rib to acquire ECG data. A Tobii X3-120 eye tracker recorded ocular data.

All 5 workstations used AVAA version 1.9.5 to access videos that resided on the Tactical Cloud Reference Implementation 1.0 “Easy” cloud. Connectivity to the cloud, which was located at the Tactical Cloud Integration Laboratory at Aberdeen Proving Ground, Maryland, occurred via the OC48 high-bandwidth network bridge. The user interface for this evaluation was the Cloud Analytics Collaboration Environment running in a Chrome 48 browser.

The full-motion color videos originated from an electro-optical and IR wide-area motion imagery system, and averaged Video National Imagery Interpretation Scale (V-NIIRS) 7+. The videos were of training events that occurred at the National Training Center at Fort Irwin, CA, and displayed no metadata overlay.

2.3 Forms and Questionnaires

For data collection we used several forms and questionnaires. We used an Internal Review Board–approved informed-consent document to provide details on participation and obtain signed consent to collect data. Our demographics form queried age, gender, formal education level, MOS (present and past), time in each MOS, time actually performing the relevant MOS duties, and other experience relevant to AVAA operations. It also captured each participant’s self-assessment of normal sleep duration, prior night’s sleep duration, and whether glasses or contacts were needed and worn.

All participants completed the NASA TLX, which captured perceived workload ratings and comparative weights for the 6 workload categories: mental demand, physical demand, temporal demand, performance, effort, and frustration. The EEG volunteers also completed a digital, modified NASA TLX for their auditory oddball task.

Participants completed the SART, subjectively rating 10 dimensions of situation awareness on a 7-point scale (1 = “less”, 7 = “more”). The dimensions were grouped according to the categories of understanding (U), demand (D), and supply (S). The understanding ratings were information quantity (UIQ1), information quality (UIQ2), and familiarity (UF). The demand ratings were instability (DI), variability (DV), and complexity (DC) of the situation. The supply ratings were arousal-readiness (SAR), spare mental capacity (SS), and concentration (SC). In accordance with the SART analysis instructions, we did not use the 10th dimension, division of attention (focus), in the overall situation awareness calculation.

A usability questionnaire captured analysts' ratings of the software's clarity, learnability, and functionality; the software's impact on the user's actions, memory, and workload; and the available user guidance and training. Ratings were "strongly agree", "agree", "neutral", "disagree", "strongly disagree", and "not applicable".

We also provided an end-of-mission report template within AVAA, which consisted of placeholders for snapshots on the first page and text boxes for responses to specific questions on the second page. We used the end-of-mission reports to assess perception via screenshots and to assess comprehension and prediction via the multiple-choice and open-ended questions.

2.4 Scenarios

We designed scenarios to provide operational context for the situation and the mission. A hard copy of the scenario was available for the analysts to reference throughout the FMV exploitation time. We did not give analysts the specific end-of-mission report questions ahead of time, but asked them to note the type of information they would typically report to their leadership. Scenario descriptions follow.

Scenario 1: Vest. Recent intelligence reporting indicates that Islamic State in Iraq and Syria (ISIS) has been training to execute some limited suicide vest events targeting the entry control points of Coalition forward operating bases in Syria. This is in direct response to ongoing US missile strikes targeting their strongholds located throughout Syria. Reporting further indicates that attack planning is being executed in the area surrounding the Al-Ramza mosque (Green), due in part to its close proximity to the safe house where foreign martyrs are being staged. Host nation security forces have identified that ISIS fighters are protecting this facility with both roving foot patrols and roof-top snipers.

Identify and annotate activity related to the rehearsal of a possible suicide vest attack. Annotate potential attack locations and suspicious activity.

The end-of-mission report questions for Scenario 1: Vest were as follows:

1) How many possible snipers can be seen moving among various rooftops?
2) What colored minaret is closest to the rubble pile?
3) What is the biggest threat to friendly troops in the area?

Scenario 2: Vehicle. During morning patrols of the local bazaar area, several military-aged males approached 2nd Battalion/7th Infantry's lead vehicle and indicated that they had overhead planning being conducted to execute vehicle-borne improvised explosive device (VBIED) events targeting Coalition convoys in and around the market area within the next 24–48 h. When asked who was planning these attacks and why they were looking to target Coalition Soldiers, the group indicated that the local ISIS leader was very upset over the recent detention of some of his family members in close proximity to area mosques. They further identified that the group was outfitting white pickup trucks with explosives in and around area mosques, and that they were planning a protest in conjunction with ISIS recruitment activities to mask the movement of these vehicles.

Annotate area mosques possibly associated with ISIS recruitment (Green/Gold) as well as trucks in close proximity to area mosques. Annotate suspicious activity, especially activity related to possible protests.

The end-of-mission report questions for Scenario 2: Vehicle were as follows:

1) How many buses are staged in close proximity to the gold minaret that could potentially be used to transport protestors on the day of the event?

2) What is the safest avenue of approach for a friendly unit deployed to monitor the protest?

3) Did you see a suspected VBIED? If so, what was the color and location?

2.5 Procedure

Before the user event, we selected videos that had a common context regarding the 2 scenarios, similar tasks, and similar target objectives. We grouped them according to their relevance and then processed them into 35 1-min segments (playlists): 17 for the Vest scenario playlist and 18 for the Vehicle scenario playlist. About half of the videos in each scenario contained scenario-related targets for identification. The other half contained reasonable targets, but they were not of specific interest in the context of the scenario. The intent was to provide realism.

A subject matter expert annotated 1–3 objects or areas of scenario interest in half of the videos in each playlist. Some of the annotations were relevant to the scenario and some were not. The other half of the videos were left with no human-made annotations. Again, the intent was to provide realism.

We then duplicated each playlist to create the 2 experiment conditions. We kept one copy as-was for the baseline condition. We processed the second copy of each playlist with the automated PVO detector algorithm so that the playlist contained a mix of computer-made and human-made annotations. The result was 2 distinct video playlists (one for each scenario), each with 2 experimental conditions applied: Vest-Baseline, Vest-PVO; Vehicle-Baseline and Vehicle-PVO. To ensure that the analysts did not collaborate, we created 3 identical copies of each of the 4 playlists so that for a given session, no 2 analysts would view the same file set. We counterbalanced the order of presentation for the baseline and PVO conditions.

We set up the user event for 5 imagery analysts per 4-h session. We had 2 sessions per day for 3 days. On Days 1 and 2 we asked for 2 volunteers for the EEG activity per session. On Day 3 we did not collect EEG data. The general schedule and approximate times follow:

1) *Informed consent.* We briefed all participants on the purpose and intent of AVAA, invited them to freely ask questions, and obtained their informed consent signature. (15 min)

2) *Calibration.* We fitted the 2 EEG volunteers with data collection sensors and equipment. Each EEG volunteer performed a 9-point calibration with a Tobii X3-120 eye-tracker to establish each individual's baseline. The setup is shown in Fig. 1 (experiment configuration). Concurrently and in a separate area we interviewed the non-EEG participants on their workflow processes. (45 min)

Fig. 1 Experiment configuration. A) The primary task monitor was used to view FMV from the AVAA software environment. B) Tobii X3-120 eye-tracker provided ocular metrics during AVAA interaction. C) A touchscreen monitor was used as a response input device during an auditory probe task as well as a digital version of the NASA TXL. D) ABM X24 EEG system was used to derive neural estimates of cognitive workload and provide auditory-evoked potentials.

3) *Training.* We provided brief training on the AVAA functionality the participants would encounter. The topics were navigation to the media and storyboard, basic search (to administratively annotate log files), play, pause, play speed, annotation chart, FMV playlist, FMV name location during play, process to add an annotation, process to edit an annotation (i.e., change the state to “confirmed”, “suggested”, or “rejected”), switch an annotation to “off screen”, take a snapshot, rename a snapshot, and open a report template to add a screenshot and type an answer. Participants had ample time to ask questions. (15 min)

4) *Practice.* A data collector sat next to each analyst, recorded notes, and ensured that he/she remained on task. Each analyst completed 2 short practice sessions to ensure understanding of the assigned tasks. In the practice sessions, participants were able to see other analysts’ annotations appear in real time as they created them. They were also able to edit them. For the sessions for record we choreographed playlist assignments so that each analyst viewed a different playlist, meaning no other analyst’s annotations would appear in a participant’s videos. (15 min)

5) *Instruction.* We read scripted instructions and the scenario out loud before the start of Scenario 1: Vest. We instructed participants that computer-generated annotations would initially appear with a yellow border, meaning the content was a “suggested” item of interest. Our human-added annotations would initially appear with a blue border, meaning the annotation was a “confirmed” item of interest. Participants were free to change the status of any annotations to “rejected” (red border), “suggested” (yellow border), or “confirmed” (blue border), if they chose to do so. We instructed the participants to watch the videos before they viewed the end-of-mission report template. (5 min)

6) *FMV exploitation.* We assigned half of the analysts to watch Vest videos in the baseline condition and half to watch them in the PVO condition. We informed them which condition they would encounter before they began. Analysts had 40 min to watch the videos and create an end-of-mission product with the provided template. We set digital timers in plain view, and at 30 min we verbally reminded everyone that 10 min remained to complete the assigned tasks. (40 min)

7) *Subjective measures.* Once participants completed the end-of-mission report, even if they finished before the allotted 40 min, they completed a NASA TLX workload rating and the SART rating. We allowed a 10-min break beginning once the last participant finished their ratings. (>10 min)

8) *Instruction.* As with Scenario 1, we read scripted instructions and the scenario aloud before the start of Scenario 2: Vehicle. We repeated the instruction that participants were to watch the videos before they viewed the end-of-mission report template. (5 min)

9) *FMV exploitation.* We assigned analysts the opposite experimental condition for the Vehicle scenario so that they experienced both baseline and PVO conditions. We informed them which condition they would encounter before they began. Analysts had 40 min to watch the videos and create an end-of-mission product with the provided template. We set digital timers in plain view, and at 30 min we verbally reminded everyone that 10 min remained to complete the assigned tasks. (40 min)

10) *Subjective measures.* Once participants completed the end-of-mission report, even if they finished before the allotted 40 min, they completed a NASA TLX workload rating and the comparison rating so that we could calculate the relative importance of each workload category for each analyst. Participants completed the usability questionnaire, provided written comments, and then verbally discussed their experiences and professional critique of the system with their respective data collectors. (15 min)

11) *After-action review.* Once all questionnaires and data collector follow-up discussions were complete we conducted an after-action review to solicit analysts' comments on AVAA from a broad implementation perspective. (10 min)

Regarding the FMV exploitation task, all participants had training and experience as an Army 35G Imagery Analyst. For the user event we asked them to use their 35G techniques and knowledge to watch the videos, annotate targets and items of interest on the video, take snapshots of targets and items of interest, change the status or details of pre-existing annotations if they so choose, and build an end-of-mission report from the provided template. Figures 2 and 3 present screenshots of the AVAA software interface and an example of a part of the end-of-mission report template.

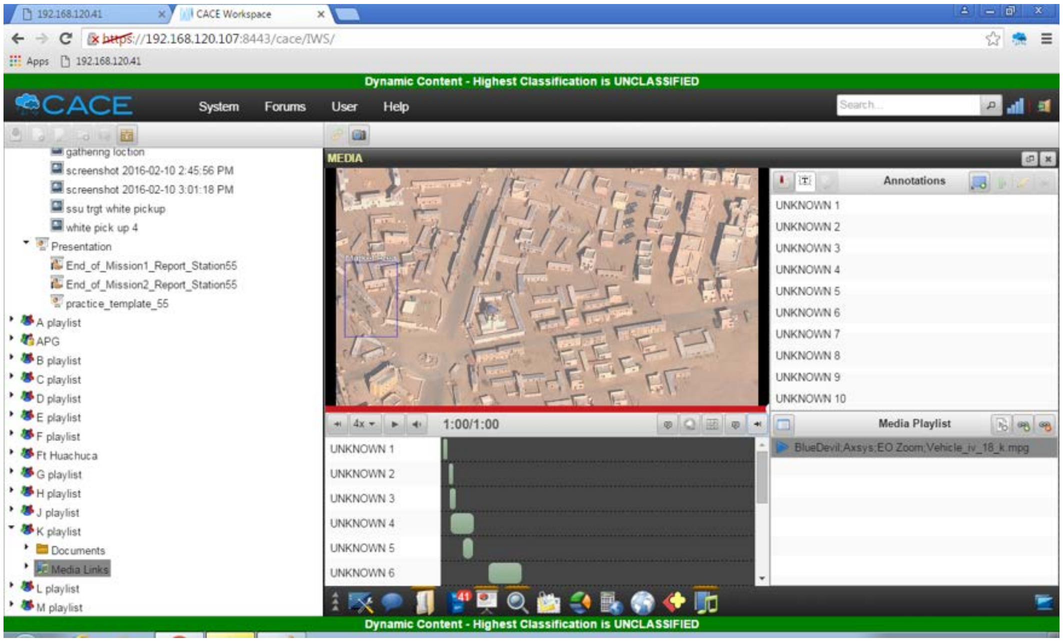

Fig. 2 AVAA software interface

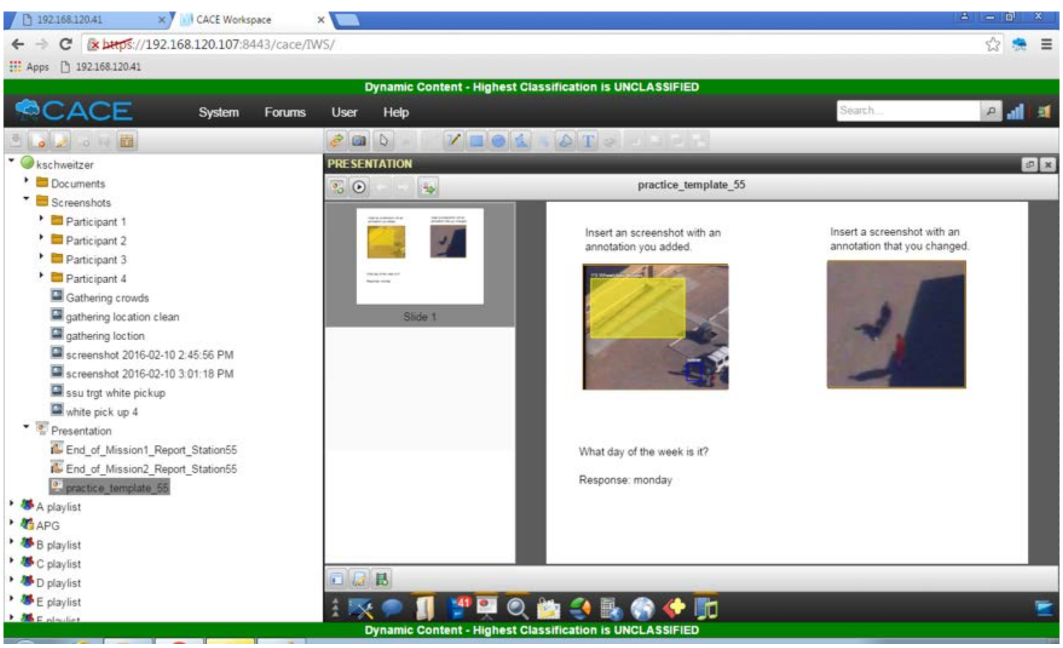

Fig. 3 End-of-mission report template

In addition to the FMV exploitation tasks, the EEG volunteers performed a simple auditory target discrimination task (2-stimulus auditory oddball) concurrently with the FMV task. This type of task has proven effective in discriminating levels of cognitive workload (Allison and Polich 2008; Miller et al. 2011). The task required participants to make a speeded response by pressing a button on a touchscreen monitor to an infrequent ($P = 0.12$, 2000 Hz) auditory stimulus presented in a series of frequent ($P = 0.88$, 1000 Hz) tones. Auditory tones were presented for 100 ms with a 3–4 s inter-stimulus interval. Event triggers associated with the sounds were sent via a split parallel port to both the external syncing unit and to the EEG and eye-tracking acquisition machine. Participants self-initiated the auditory task by pressing a button on the touchscreen immediately after they were given their primary task instructions. Participants stopped the auditory task using the touchscreen after completing their FMV exploitation and end-of-mission report. Starting and stopping the auditory task also generated unique event codes though the split parallel port in order to synchronize the EEG and eye-tracking data based on the target search task condition. After participants found their intended target and stopped the auditory task, they completed a digital version of the NASA TLX using a 100-point visual analogue scale. Due to the nature of the task, the physical demand factor on the NASA TLX was not included. For the eye-tracker, we sampled data from each eye at 120 Hz using custom software with the Tobii SDK.

3. Results and Discussion

We used a multi-aspect approach to assess operator functional state as a means to evaluate the AVAA system design. This approach may be applied to a number of user states in a variety of computer workstation environments, but here we focused specifically on cognitive workload during FMV analysis.

We recorded the actual amount of time (in minutes) participants took to complete the scenario ($m_{Baseline} = 40.4$, $SD_{Baseline} = 15.9$; $m_{PVO} = 44.5$, $SD_{PVO} = 20.5$). No significant differences were expected due to the study-imposed time limit of 40 min. We allowed participants to quickly finish the task they were working at the 40-min mark, so some total times slightly exceeded 40 min. A 2-tailed paired t-test found no significant differences between the Baseline and PVO conditions for total scenario time ($t(26) = 0.88$, $p = 0.39$).

Log files captured all button clicks in comma-separated values files. From those files we tallied the number of annotations added ($m_{Baseline} = 14.1$, $SD_{Baseline} = 9.0$; $m_{PVO} = 12.6$, $SD_{PVO} = 8.3$) and number of annotations edited ($m_{Baseline} = 37.1$, $SD_{Baseline} = 61.5$; $m_{PVO} = 27.4$, $SD_{PVO} = 35.3$). A 2-tailed paired t-test found no significant differences between the Baseline and PVO conditions for annotations added ($t(26) = 1.29$, $p = 0.21$) or for annotations edited ($t(26) = 1.02$, $p = 0.32$).

Participants were allowed to change annotation labels to or from “rejected”, “suggested”, or “confirmed”. Analysts informed us these labels were incompatible with their duty lexicon, where a declaration of “confirmed” intelligence requires far more evidence than we provided during the exercise. As a result, some participants chose to spend time changing many “confirmed” annotations to “suggested” or “rejected”.

We also tallied the number of clicks on the play-pause soft button ($m_{Baseline} = 135.2$, $SD_{Baseline} = 82.9$; $m_{PVO} = 123.7$, $SD_{PVO} = 51.1$). A 2-tailed paired t-test found no significant differences between the Baseline and PVO conditions for number of play–pause soft button clicks ($t(26) =0 .86$, $p = 0.40$).

We tallied and compared the number of snapshots taken ($m_{Baseline} = 2.9$, $SD_{Baseline} = 1.6$; $m_{PVO} = 3.1$, $SD_{PVO} = 1.3$) and the number of snapshots annotated ($m_{Baseline} = 1.9$, $SD_{Baseline} = 1.6$; $m_{PVO} = 2.2$, $SD_{PVO} = 1.4$). A 2-tailed paired t-test found no significant differences between the Baseline and PVO conditions for total number of snapshots taken ($t(22) = 0.59$, $p = 0.56$) and number of snapshots annotated ($t(22) = 0.79$, $p = 0.44$).

We graded the end-of-mission report for accuracy and thoroughness ($m_{Baseline} = 2.7$, $SD_{Baseline} = 1.8$; $m_{PVO} = 2.5$, $SD_{PVO} = 2.1$). For Scenario 1 reports, scoring for #1 and #2 was 2 points for a correct answer, 1 point for a partially correct answer, and 0 points for an incorrect answer. Scoring for #3 was 2 points for an answer with justification, 1 point for an answer with no justification, and 0 points for no answer. For Scenario 2 end-of-mission reports, scoring for #1 was 2 points for a correct answer, 1 point for a partially correct answer, and 0 points for an incorrect answer. Scoring for #2 and #3 was 2 points for an answer with justification, 1 point for an answer with no justification, and 0 points for no answer. A 2-tailed paired t-test identified no significant differences between the Baseline and PVO conditions ($t(20) = 0.29$, $p =0.76$).

For the subjective measures, we calculated weighted cognitive workload ratings ($m_{Baseline} = 9.9$, $SD_{Baseline} = 2.8$; $m_{PVO} = 9.5$, $SD_{PVO} = 2.9$) for the NASA TLX data according to Hart and Staveland (1988). A 2-tailed paired t-test identified no significant differences between the Baseline and PVO conditions ($t(25) = 1.02$, $p = 0.32$).

We calculated the overall perceived situation awareness rating for the SART, as the average U ratings minus the average D ratings less the average S ratings (i.e., U – [D – S]) were U = [UIQ1 + UIQ2 + UF]/3, D = [DI + DV + DC]/3, and S = [SAR + SS + SC]/3). A 2-tailed paired t-test found no significant differences between the Baseline and PVO conditions ($t(26) = 0.303$, $p = 0.76$).

3.1 Usability and User Comments

The focus for the Phase II evaluation was functionality, not necessarily usability, as the interface through which users accessed the AVAA pipelines was not the final construct that AVAA will use. However, we did examine certain usability points that were applicable to necessary functionality. Table 1 shows the tallied ratings for the functionality-related questions. A full listing of the usability questionnaire responses and user comments is available in the Appendix.

Table 1 Usability questionnaire responses related to functionality

Statement	Agree + strongly agree	Neutral	Disagree + strongly disagree	Not applicable
AVAA provides all the information I need to do my work.	8	6	11	2
I can understand and act on the information provided.	24	0	3	0
AVAA directs my attention to critical or abnormal data.	13	10	4	0
Learning to use this software is easy.	25	2	0	0
I feel confident in my ability to complete my assigned task using AVAA.	20	6	1	0
Compared with my current method of exploiting imagery, AVAA does not affect my workload.[a]	15	6	4	2
Compared with my current method of exploiting imagery, AVAA decreases my workload.[a]	10	8	6	3

[a]These are the subjective user responses to the primary question we scientifically addressed in this study.

The user comments were freely offered assessments and opinions from the analysts. The frequency of mentions, however, only indicated how many analysts thought to verbalize a point. This is not to imply that analysts who failed to mention the same points thought them unimportant.

The frequently cited functionality comments that participants provided centered heavily on needing directional information on the video and having a time stamp and/or metadata reference on the video. Analysts considered the availability of a geo-rectified map database and the ability to export annotations that were compatible with standard map products to be important.

Analysts cited a need for pinned annotations, meaning the annotation is fixed to either a geographical point or the pixels of the moving object; as it was, manual tracking distracted from the mission. Some participants found the computer-generated annotations either too distracting or they wanted to see more-meaningful descriptors. Several wanted more options for annotation shapes and orientations or did not want the shapes to have a distracting color fill upon selection.

Analysts noted that AVAA needs to use the Army's standard language to describe targets as "possible" (25% certainty), "suspected" (50% certainty), "probable" (75% certainty), and "confirmed" (100% certainty). The analysts in this study stated they never "confirm" a target at their level, as confirmation requires 3 sources essentially with eyes-on the target.

Several comments involved video controls, which would allow the participants to analyze the videos more effectively. Suggestions included frame-by-frame video stepping, reverse play, and automatically stopping the video when adding an annotation.

Another notable issue was the need to sync, when possible, AVAA hot-key functionality with what analysts are accustomed to using. Several participants commented on making mistakes out of habit due to different hot keys and mouse clicks.

Analysts provided a healthy list of recommendations that would improve the user interface of AVAA in general, and those are consolidated in the Appendix.

3.2 Volunteers for Physiological Measures

For the subset of EEG volunteers (n = 8), we evaluated both continuous and discrete electrophysiological estimates of cognitive workload. Additionally, we collected ocular metrics, ECG data, behavioral responses to a secondary task, and questionnaire data from a modified NASA TLX. This approach allowed us to obtain operator functional-state data during 2 modes of operation using a multimeasurement system. It was hypothesized that the 2 modes of operation would result in significantly different workload levels. Metrics included heart-rate variability, ratio of theta power at electrode Fz over alpha power at electrode Pz, N100 auditory-evoked potential amplitude, and blink duration.

Workload metrics were estimated from the EEG by calculating the ratio of power in the theta frequency (3–7 Hz) at electrode Fz over the power in the alpha frequency (8–12 Hz) at electrode Pz. Workload estimates were derived once per second. Power spectral density values were computed by performing a fast Fourier transform and applying a 50% overlapping window across 1-s epochs in 1-Hz bins from 1 to 40 Hz.

Epochs were extracted from the EEG (–200 to 1000 ms) and time-locked to the onset of the auditory stimuli. Epochs were averaged to create ERPs for both standard and target-oddball stimuli. This was done for each target search mission within each condition (Baseline and PVO). While ERPs were generated for both the target and standard auditory stimuli, the target stimuli presented in the auditory task were primarily used as a behavioral performance metric. We focused on the ERPs from the frequent standard stimuli, as they provided more samples over time, which allowed us to investigate how the amplitude of this potential changed as a function of condition as well as their task engagement, indicated by their behavioral responses to the auditory targets. Here, task engagement was defined as periods of time in which the user failed to respond to the auditory targets. It was expected that the ERPs to the auditory stimuli would change in amplitude between the Baseline and PVO conditions if one of the conditions induced more cognitive demands on the operator.

The mean amplitude (150–175 ms) of the N1 auditory-evoked potential (AEP) in the Baseline ($m_{Baseline}$ = –4.52, $SD_{Baseline}$ = 2.92) and PVO (m_{PVO} = –4.62, SD_{PVO} = 2.29) conditions was compared using a one-way analysis of variance (ANOVA). No significant differences were found ($F(1,7) = 0.56$, $p = 0.82$), indicating that the N1 AEP amplitude was on average similarly affected in the Baseline and PVO conditions (Fig. 4).

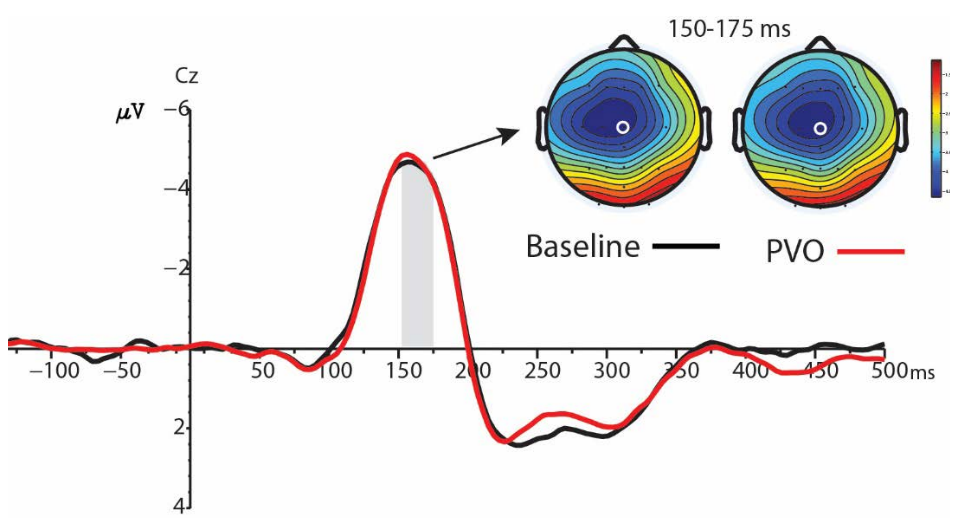

Fig. 4 Auditory evoked potentials showing the N1 component of all standard tones taken from electrode Cz in the Baseline and PVO conditions. The topographical voltage maps show the average voltage distribution between the 2 conditions taken 150–175 ms after stimulus onset.

Accuracy and reaction time for the auditory targets in the secondary task were compared between the Baseline and PVO conditions. The results indicated that operators performed equally well in the 2 conditions, as no significant difference was obtained for accuracy ($F(1,7) = 1.61$, $p = 0.25$) or reaction time ($F(1,7) = 0.013$, $p = 0.91$) (Fig. 5A and 5B).

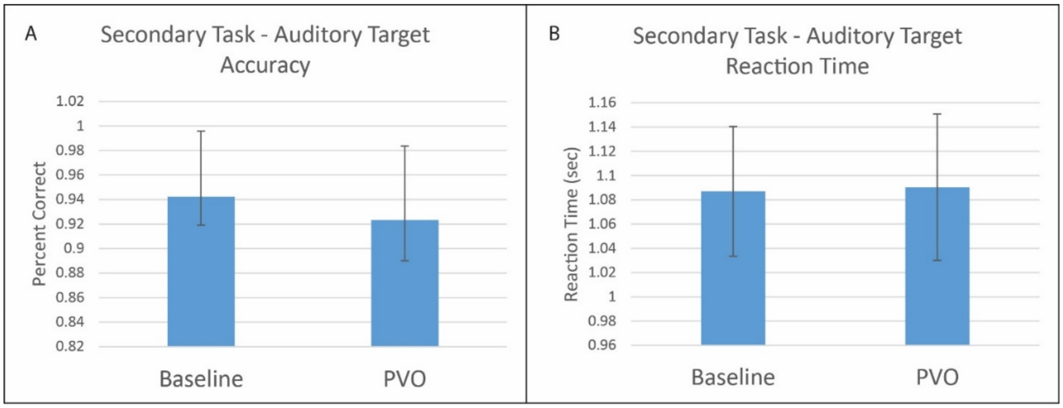

Fig. 5 Average behavioral performance for the secondary auditory oddball task. A) Percent correct responses to auditory targets. B) Reaction time in seconds to auditory targets.

Heart-rate variability was calculated for the Baseline and PVO conditions using the SD of heart rate in beats per minute (Fig. 6A). No significant differences were found between these conditions ($m_{Baseline} = 44.4$, $SD_{Baseline} = 12.89$; $m_{PVO} = 46.35$, $SD_{PVO} = 15.69$) using a one-way ANOVA ($F(1,7) = 0.34$, $p = 0.58$).

The EEG-derived workload estimate was analyzed using a one-way ANOVA. The results showed this measure was not significantly different ($F(1,7) = 1.64$, $p = 0.24$) between the Baseline ($m_{Baseline} = 1.28$, $SD_{Baseline} = 0.087$) and PVO ($m_{PVO} = 1.27$, $SD_{PVO} = 0.089$) conditions, indicating that the average estimated workload was similar between the 2 conditions (Fig. 6B).

We used eye-tracking data to measure saccade and fixation metrics as well as provide estimates of gaze distribution. We calculated eye fixations using the algorithm described by Engbert and Mergenthaler (2006). Saccade magnitude and blink duration were evaluated in the Baseline and PVO conditions using the same one-way ANOVA as before. The results indicated no significant differences between the conditions in either saccade magnitude ($F(1,7) = 0.49$, $p = 0.51$) or blink duration ($F(1,7) = 2.65$, $p = 0.15$) (Fig. 6C and 6D).

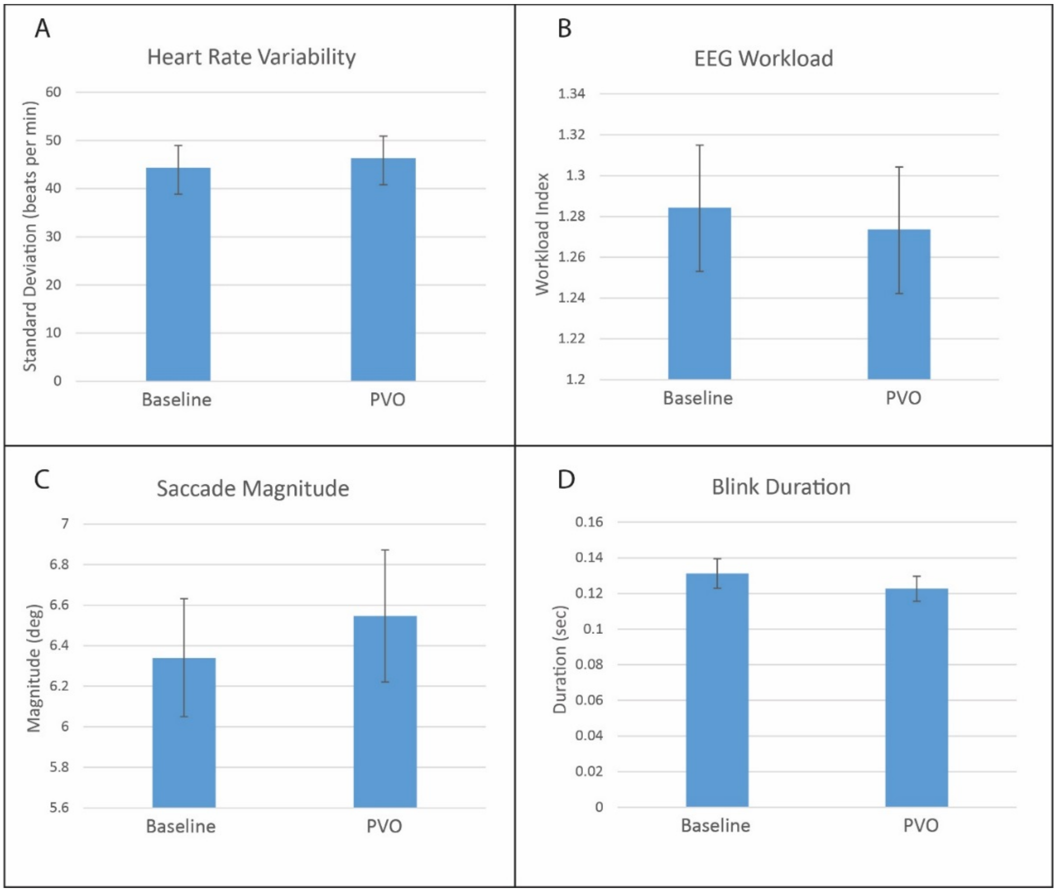

Fig. 6 ECG, EEG, and ocular-based metrics in the Baseline and PVO conditions. A) Heart-rate variability calculated as the SD of heart rate in beats per minute. B) Workload estimates depicting the ratio of power in the theta frequency taken from electrode Fz over the power in the alpha frequency obtained from electrode Pz. C) Average saccade size in visual angle. D) Average blink duration in seconds. Error bars = ±standard error.

Figure 7 shows the distribution of fixations between the Baseline and PVO conditions during the first and last 10 min of the task for participant S04. These data show that many fixations were distributed around the video window, with many focused on the video controls especially in the first 10 min. The pattern of fixations changed in the last 10 min of the task as more fixations were directed to left side of the AVAA screen with fewer fixations on the video controls.

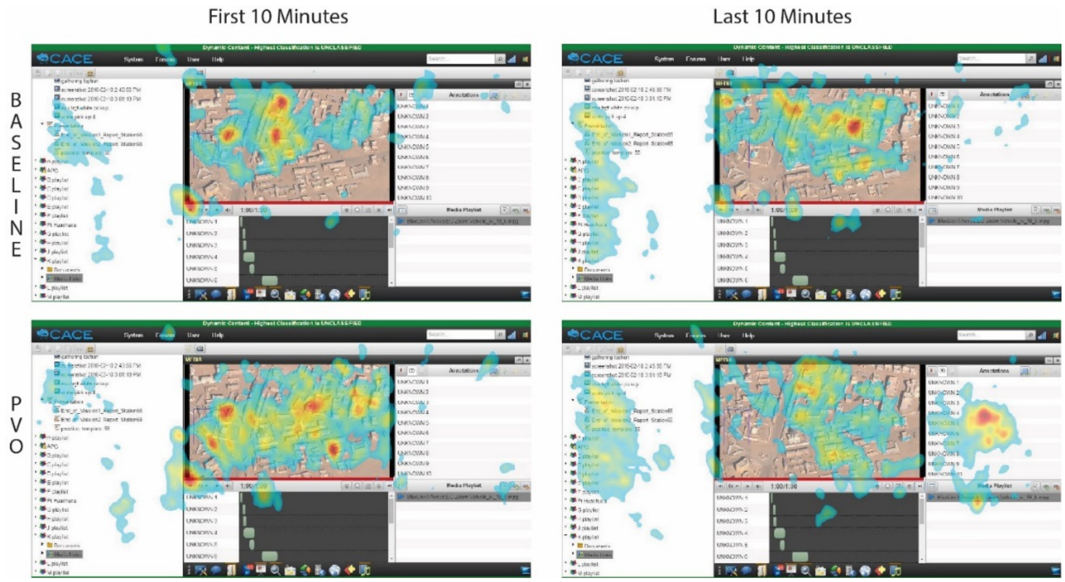

Fig. 7 Eye-fixation distribution during the first and last 10 min of the task in each condition for participant S04. The video frame depicted is for illustrative purposes only.

The primary advantage of using EEG and eye-tracking together with behavior is to provide a continuous, multi-aspect estimate of cognitive state over time. This allows evaluators the ability to identify periods of fluctuations in the data used to estimate cognitive state and relate them to user and system performance. Figure 8 shows normalized continuous data obtained from multiple measures during the first mission for S04. The figure shows multiple fluctuations in the measures over time and highlights how subtleties in data can be lost in averages. One striking observation is the strong relationship between pupil diameter and the EEG workload metric derived from the theta and alpha power ratio. The cyan diamonds show single-trial-evoked neural responses at electrode Cz from each auditory nontarget stimulus. The blue circles show the reaction time to the auditory target stimulus. Together these data show how multiple continuous measures (ocular, neural, and behavioral) can be used to identify changes in cognitive state that may be masked in aggregated responses.

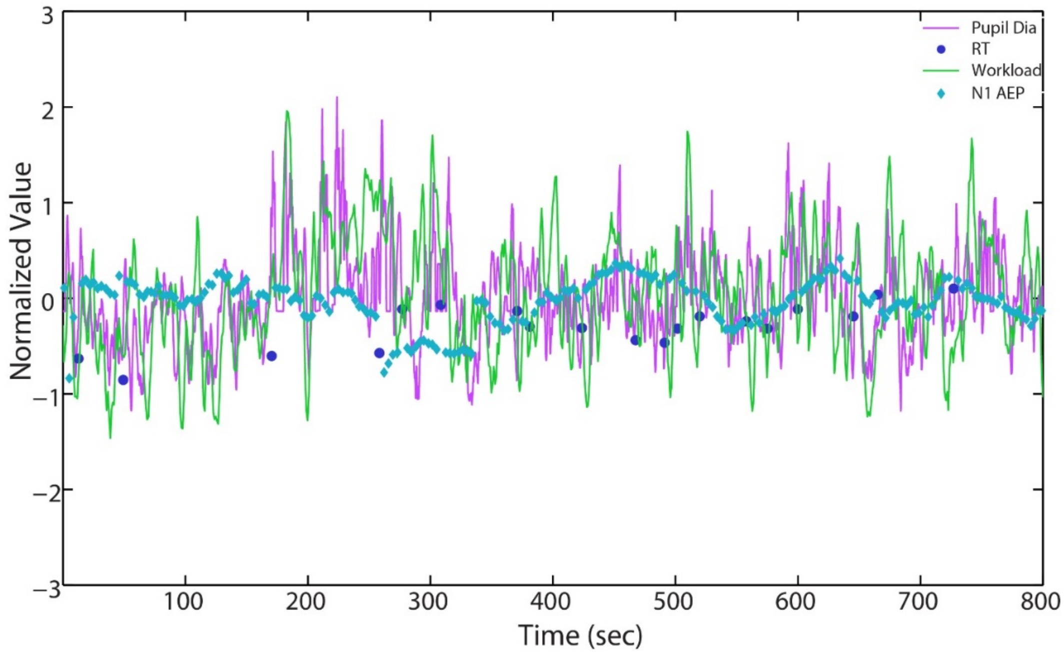

Fig. 8 Continuous ocular, neural, and behavioral measures: normalized values from pupil diameter (purple line), EEG-derived workload estimates (green line), single-trial-evoked response from nontarget auditory stimuli (cyan diamonds), and reaction time to auditory targets (blue circles)

On average, none of the electrophysiological, ocular, or behavior indices of workload were significantly different between the Baseline and PVO conditions. While differences in workload may have existed between these conditions, it is difficult to ascertain from average measures over the extended time period that operators were engaged with the AVAA task. We present a continuous representation of how workload may have fluctuated over time using pupil, neural-based, and behavioral measures. The failure to find significant differences using the physiological measures agreed with NASA-TLX subjective measures, thus increasing our confidence that the workload difference for the 2 experiment conditions was not an important factor in this experiment. In future experiments, directly correlating fluctuations in these continuous measures with both user performance in the primary task as well as the state of the system may provide meaningful insight into user/system interactions.

3.3 Limitations and Caveats

The Phase II data collection was scheduled sooner than system readiness might have dictated, due in part to contract end dates. AVAA system engineers had only the PVO pipeline ready for evaluation, and in the hurry to have everything operational, they inadvertently installed an older version of the PVO software. As a consequence, participants saw a somewhat simpler PVO annotation capability than what the program initially advertised.

Also, due to the constraints of the cloud connectivity at Fort Huachuca, AVAA did not have access to geo-rectified maps at the time of the user event. We explained this to participants, with the understanding that AVAA is capable of displaying the maps but was unable to demonstrate it at the time of data collection.

4. Conclusions and Discussion

The focus of the Phase II human factors functionality evaluation was to determine whether AVAA's CV algorithm pipeline for an automated PVO detector impacted analyst workload and performance during realistic scenario-based operations. The results from multiple measures indicate that AVAA's PVO pipeline did not significantly affect the measures we analyzed for performance, cognitive workload, or situation awareness as compared with the baseline AVAA system. While it is always risky to base conclusions on failures to reject the null hypothesis, the findings of no real differences using a variety of measurement techniques does suggest that there was little additional benefit to adding the PVO CV to the analyst's workstation in the current experiment. However, some observers pointed out that the utility of the PVO was not highlighted in our experimental paradigm. The experienced 35G analysts were able to extract sufficient data from the non-PVO videos to conduct their intelligence processing. In many real-world environments, the number of videos to be sampled could be in the hundreds to thousands. A filter that enabled analysts to examine a subset that contained people, vehicles, or objects would useful for many mission environments. In our previous experiments, filtering using VNIRS technology reduced the number of videos and processing times experienced analysts required to conduct their mission (McDermott et al. 2015). AVAA is still under development; additional CV algorithms and advanced interfaces that permit further sorting of CV products are being investigated. A more-mature AVAA or similar analytic and visualization tools will be important technologies for evaluating and filtering FMVs during antiterrorist, police, and military operations.

Analysts identified a critical need for directional markers and time stamps and insisted that software labels must be compliant with user lexicon. Analysts identified multiple usability recommendations for AVAA. While outside the scope of this study, the information is extremely useful and was thus compiled in the Appendix. A related report evaluated AVAA by modeling its impact on the cognitive workload of imagery analysts (Plott et al. 2017).

References

Ahlstrom U, Friedman-Berg FJ. Using eye movement activity as a correlate of cognitive workload. Int J Ind Ergo. 2006;36(7):623–636.

Allison BZ, Polich J. Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. Bio Psyc. 2008;77(3):277–283.

Backs RW, Walrath LC. Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. Appl Ergo. 1992;23(4):243–254.

Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, Zivkovic VT, Craven PL. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation Space Enviro Med. 2007;78(5):B231–B244.

Brookhuis KA, de Waard D. Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis and Prevention. 2010;42(3):898–903.

Brookings JB, Wilson GF, Swain CR. Psychophysiological responses to changes in workload during simulated air traffic control. Bio Psyc. 1996;42(3):361–377.

Brouwer AM, Hogervorst MA, van Erp JBF, Heffelaar T, Zimmerman PH, Oostenveld R. Estimating workload using EEG spectral power and ERPs in the n-back task. J Neural Eng. 2012;9(4):045008.

Dinges DF, Mallis MM, Maislin G, Powell I. Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. Washington (DC): National Highway Traffic Safety Administration; 1998. Report No.: HS-808 762.

Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Meth Inst Comp. 1985;17(6):652–655.

Engbert R, Mergenthaler K. Microsaccades are triggered by low retinal image slip. Proceedings of the National Academy of Sciences. 2006;103(18):7192–7197.

Fairclough SH, Venables L, Tattersall A. The influence of task demand and learning on the psychophysiological response. Int J Psychophys. 2005;56(2):171–184.

Fournier LR, Wilson GF, Swain CR. Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. Int J Psychophys. 1999;31(2):129–145.

Gevins A, Smith ME. Neurophysiological measures of cognitive workload during human-computer interaction. Theor Iss Ergo Sci. 2003;4(1-2):113–131.

Gevins A, Smith ME, Leong H, McEvoy L, Whitfield S, Du R, Rush G. Monitoring working memory load during computer-based tasks with EEG pattern rcognition methods. Human Factors and Ergonomics Society HFES. 1998;40(1):79–91.

Gundel A, Wilson GF. Topographical changes in the ongoing EEG related to the difficulty of mental tasks. Brain Topog. 1992;5(1):17–25.

Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Advances in psychology. 1988;52:139–183.

Johnson RR, Popovic DP, Olmstead RE, Stikic M, Levendowski DJ, Berka C. Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. Bio Psych. 2011;87(2):241–250.

Kramer AF, Trejo LJ, Humphrey D. Assessment of mental workload with task-irrelevant auditory probes. Bio Psych. 1995;40(1–2):83–100.

Makeig S, Inlow M. Lapse in alertness: coherence of fluctuations in performance and EEG spectrum. Electronic Component News. 1993;86(1):23–35.

McDermott PL, Plott BM, Ries AJ, Touryan J, Barnes MJ, Schweitzer KM. Advanced video activity analytics (AVAA): human factors evaluation. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2015. Report No.: ARL-TR-7286.

Miller MW, Rietschel JC, McDonald CG, Hatfield BD. A novel approach to the physiological measurement of mental workload. Int J Psychophys. 2011;80(1):75–78.

Moroney WF, Biers DW, Eggemeier FT. Some measurement and methodological considerations in the application of subjective workload measurement techniques. Int J Aviation Psych. 1995;5(1):87–106.

Mulder B, de Waard D, Brookhuis KA. Estimating mental effort using heart rate and heart rate variability. In: Stanton N Hedge A, Brookhuis K, Salas E, Hendrick HW, editors. Handbook of ergonomics and human factors methods. Boca Raton (FL): CRC Press.; 2004.

O'Donnell RD, Eggemeier TF. Workload assessment methodology. In: Boff KR, Kaufman L, Thomas JP, editors. Handbook of perception and human performance, vol. 2: cognitive processes and performance. Oxford, (England): John Wiley & Sons; 1986. p. 1–49.

Plott BM, McDermott PL, Barnes MJ. Advanced Video Activity Analytics (AVAA) human performance model report. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2017.

Stikic M, Johnson RR, Levendowski DJ, Popovic DP, Olmstead RE, Berka C. EEG-derived estimators of present and future cognitive performance. Frontiers in Human Neuroscience. 2011;5.

Van Orden KF, Limbert W, Makeig S, Jung TP. Eye activity correlates of workload during a visuospatial memory task. Human Factors and Ergonomic Society HFES. 2001;43(1):111–121.

Wilson GF. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int J Aviation Psych. 2002;12(1):3–18.

INTENTIONALLY LEFT BLANK.

Appendix. Usability Ratings and User Comments

Usability Questionnaire Responses	Agree + Strongly Agree	Neutral	Disagree + Strongly Disagree	Not Applicable
1. AVAA provides all the information I need to do my work.	8	6	11	2
2. I can understand and act on the information provided.	24	0	3	0
3. The resulting operations of the numeric, function, and control keys are the same as for other tasks.	14	6	3	4
4. AVAA directs my attention to critical or abnormal data.	13	10	4	0
5. When a keystroke (or mouse click) does not immediately produce the response I expect, the software gives me a message, symbol, or sign to acknowledge my input.	4	8	10	5
6. Whenever I am about to enter a critical change or take some important, unrecoverable action, I must confirm the entry before accepting it.	15	2	4	6
7. I can backtrack to the previous menu by using a single keystroke or mouse click.	16	3	1	7
8. The walk-through training gave me sufficient guidance so that I was able to complete my assigned task.	24	2	1	0
9. Learning to use this software is easy.	25	2	0	0
10. I feel confident in my ability to complete my assigned task using AVAA.	20	6	1	0
11. Compared to my current method of exploiting imagery, AVAA does not affect my workload.	15	6	4	2
12. Compared to my current method of exploiting imagery, AVAA decreases my workload.	10	8	6	3

User Comments—What would you do to improve the AVAA system?

<u>Orientation</u>

- Cardinal direction (heads up display) with platform and threat location
- Have MGRS on screen
- Annotations should possibly have MGRS added to them
- No GEO/MGRS

- Grid values
- At a minimum, need directional orientation and meta data
- Coordinates
- Need crosshair with coordinate.
- Directional values
- Direction - north arrow in mission
- North arrow
- North arrow
- Get a north arrow present
- North arrows
- Also a north arrow
- The video definitely needs a north arrow on video feed.
- Maybe add a north arrow to the video segments
- Geo-rectified database is accessible via DCGS-A
- Automatic recognition for georectified location, i.e. know location mosque, churches, capital buildings
- Running AVAA with Google Earth running in the background
- Integrated map
- Map feature loaded or have map reference for listing.

<u>Time</u>

- Time stamp aligned with video stream times
- Time stamp
- There was no time stamps for annotations.
- Time stamp in titles, especially because they were not in order
- Timeline for videos - chronological order is CRITICAL. Need time stamps, at least.
- Video should be sequential

- Time bar in mission
- No time bar
- Not having video chronology is really bad - it hampered ability to build intel.

Video player functionality

- Playback buttons
- Replay-in-reverse option
- More hot keys for FF, RW, speed
- Playlist will not play next video automatically.
- Have multiple video feeds able to be played at once.
- Scroll on play/time bar to re-watch repeatedly instead of clicks to control video
- Step through frame by frame.
- Want to step frame by frame (or at least slower stepping)
- Add a zoom wheel
- Zoom in/out

Video

- Polarity of video feed
- Provide metadata for the sensor.
- At a minimum, need directional orientation and meta data
- Must have meta data!!
- HUD (UAV info)
- No HUD, metadata
- Add sensor selection.

Annotations and snapshots

- Too many false positives (from pro); won't reduce the workload
- System annotations were annoying, all over the place.
- System-generated annotations were clutter. If only a couple, he would check them, otherwise they are annoying and basically ignored.
- Algorithms that detect activity need to improve.
- Want more background on what causes the algorithm to key on a target.
- When you drop box on an object, it should lock on that object.
- Steady annotations that lock on target.
- ID and lock on would be good
- The ability to fix an annotation to a pixel signature (i.e. a person on screen)
- User-generated annotations were great, but should only persist a few seconds due to view drift.
- Allow annotation to lock to item not the whole screen.
- Want annotations to track coordinates.
- Spatially related annotations with visible coordinate system
- Georegistered annotations
- Would like to see annotation stay on geographic location.
- Annotations that stay attached to the ground
- Convert graphics to ground space.
- More intuitive annotations. It felt as though the analyst had to remember to let the annotation know when the object was no longer in scene.
- Change the "state" in the create annotations to a standard "possible, probable, confirmed" instead of "confirmed, suspected, rejected".
- "Confirmation" has a very specific meaning for Analysts
- Analysts do not "confirm" anything without 110% certainty.
- Possible (25%), Suspected (50%), Probable (75%), Confirmed (100%; need 3 sources to confirm… analysts at his level NEVER confirm!)
- Annotations --> move to army standard (poss/prob/conf)

- Work on the annotation queueing area working more effectively. Multiple videos the area didn't work.
- Annotations are not very user friendly.
- Annotations annoying and distracting. Building 10, man 11, but no man there.
- Paid more attention to blue annotations because they had more detail; yellow box with "unknown" was more of a distraction.
- Paid more attention because of the descriptors, not the outlines themselves.
- Thought yellow annotations were way off - did not trust labels
- Normally capture full video screen for snapshots
- Snapshot of entire screen, not selection
- 2 options for screenshots - full screen and customizable

Analysis products

- Product building compatibility
- Product creation
- Annotations should possibly be exportable as shapefile option.
- Option to load/edit templates on screen and be able to export or save as a jpeg/pdf
- Would like to see features that allow for video chipping pushed to PowerPoint production.
- Presentation mode should reflect PowerPoint better.
- Fix the presentation develop issue
- Ability to use actual overlays templates for products (think SOCET GXP)
- I would add a template option
- Definitely want to be able to import a template.
- ID numbers, who made it, map chip, etc.
- Relabeling of clip names for more clarity

Visualization

- Want more screen real estate for video because it is the major task.
- Ability to resize windows
- Setup a little better to be able to have the multiple facets up while viewing a feed.
- Box to side with metadata instead of top and bottom of image so don't look away from image. Perhaps have as preference.
- Annotations should never be filled
- Does not like color fill when annotation is selected.
- Annotation bar - being able to drag one direction or other (drag it to location).
- Automatically stopping the feed when adding annotation.
- As soon as user selects annotation tool, auto-pause the video.
- Additional window to type in while screening
- Be able to set time limits for graphics, i.e. show for 5 seconds, then disappear.
- Want more ways to annotate differently.
- Want different shapes and orientations for annotation borders.
- Have more annotation options.
- Annotations are too generic (e.g. circle, point)
- Liked being able to enter full comments on annotations.
- When enter first date box, want second date box to default to same month/date and move from there.
- Screenshots should be instant rename
- Toggle functionality to display either only man or machine generated annotations.
- The system-generated annotations were keying on important things, but need to be able to turn them off.
- "[T]" will hide all annotations

- Want drag-drop for window apps.
- SOCKET, PPTX, etc.
- Fix the layers

Overall

- This is a big improvement over earlier DCGS-A effort, which was a multi-server system.
- Collaboration - only one would actually annotate - good to see!
- Collaboration amongst users would be GOOD.
- Sufficient functionality
- As far as software goes it is user friendly.
- Great program and extremely user friendly.
- Great new software. I am interested in learning more about AVAA.
- Usability was easy
- User-friendly software base
- Pretty easy
- Interesting
- Strategy did not change with AVAA.
- The exploitation and familiarization will take additional training.
- I don't see a difference in what I use already except this is very glitchy.
- System seemed to do better with vehicles.
- Would need more time using system.
- Cannot comment due to full software is not loaded with map.
- I had no clue where I am looking and at what time.
- Lag on text box response
- Words should be clearer
- No slashes "/" in file names for screenshots

- Kept forgetting right click to rename instead of double click so toggled between video and snapshot more than desired.
- More shortcut keys
- Hot keys are extremely helpful with speeding up exploitation.
- Small fixes to streamline for users
- Some of the anomalies are a bit redundant but I can ignore them.
- Once AVAA is refined, PVO might be great for rural, but not expected to be useful for city work.
- Exploitation of real time FMV
- Helps to get immediate feedback into report while focused on mission
- System-generated annotations were extremely functional, but can be cluttering. It was catching things like antennas right away.
- Too many system-generated annotations; would rather see none.
- System flags too many false boxes - mostly annoyance

<u>Study-related</u>

- No mission set to go by -- the scenario was not relevant except roving patrols.
- Need more info on scenario ex. MGRS of FOB and suspected safe house
- Training --> show a completed product 1st.

Additional User Comments

To be useful, how would look like, behave?

- Let user be able to pick parameters what looking for.

How helpful was the tool?

- Wish had it with FMV. Streamlines ability to make product incredibly user friendly.

- Side view map of where are in the country.
- Infrared would be helpful in Scenario 2. Sensor switchery gives analyst some idea.

During the mission phases do you feel that the use of AVAA makes your job easier, harder, or has no impact? Why?

- It makes the job harder at the moment because of product development is almost non-existent.
- Harder. User-gen annotations didn't have intel value. Couldn't confirm where is sign that said that.
- No impact
- Yes - like layout, switch between tasks; good organization
- "ouerzu" yes. Need metadata or geo/map info.
- Compares to MAWS
- Seemed to be easier
- Ease of use; "hot" keys
- Live mission - annotations

Specifically for the post mission reporting do you feel that the use of AVAA makes preparing reports easier, harder, or has no impact? Why?

- Harder. Extremely harder with no directional value or time scope. Also no way to successfully build a product. The reporting is impossible.
- Little or no value; current tools also do this.
- Easier - tool easy to modify
- Text edits stinks. Hard to discern "mode".
- Being able to upload templates and download jpegs and mpegs.
- Easier. Ability to plug in pre-formatted products to build end-of-mission products.
- Yes, easy enough to navigate the production of reports

Are there circumstances or missions in which AVAA would be useful? Would not be useful?

- Real time FMV could be useful with AVAA.
- Right now it's not the best for forensics.
- None
- Forensic good/prime to annotate
- Share with multiple people
- Same as MAWS
- Could be useful in real time reporting and forensics
- Yes
- No. It's too broad.

What are your impressions of the AVAA imagery viewing functions (play, pause, rewind, fast forward, jump-to, back-to-real-time)?

- Once again non-existent. It plays pauses and goes fast but not backwards at all.
- User friendly, simple
- Easy to use; smooth operation
- Pretty good; annotations list is too large; would like to see more "lineao" annotations list.
- OK
- Well received; easy to use
- No overkill is appreciated.
- Like interface; easy to use

Did AVAA support time-mode recognition / management (Am I looking at live or in the past? Where am I in the past?)

- No. NEEDS DTG [date time group] for reference.

- Anything else noticed? Can exploit in another. If can do products --> useful.
- Yes
- Good; yes - maneuver around "grid"
- No
- Yes, but map services are a must!

Did you notice any data transmission delays, intermittency, or loss notification / management / documentation with AVAA?

- No
- No
- Nope
- Yes (VAWS went out.)
- No
- Yes, during presentation and using "control/A" function from screen
- Only when dragging slider

If so, did it impact your mission performance?

- No
- What like AVAA in general? Potential for useful. Did go frame-frame with arrows. Easy to use, needs more. Can't save MGRS to drag + drop. Can't make GEOINT product without imparting save where else.
- Halted for a bit
- Reset system
- Slower

What aspects of AVAA help you do your job?

- None at this time
- If was in FMV unit - yes.

- Simplified forensics
- Pretty good resolution. High res video helps to identify equipment, people (gender), etc.; zoom level added clarity.
- Snapshots - easy and right there.

What aspects of AVAA hinder you from doing your job?

- No directional reference.
- No time reference (in mission and time of day).
- Manual tracking distracts from mission.
- No
- None
- Addition of -N- bearing required!
- None
- If map is not in final product.
- Distracting machine annotations

There are advanced features of AVAA under development. I'd like to describe them and get your impression of them and under what circumstances they would be useful. [PVO detector, PVO tracker, license place tracker, face detector/tracker]

- PVO tracker - useful, especially because hard to keep camera steady.
- License plate tracker - never had LP in theater, usually overhead view, easily corrupted. 9/10 things happen at night so color irrelevant (i.e. red truck). Most things over there are black or whole colored.
- Face detector/tracker - how good imagery needed - UAVs don't have capability - if not 90% confidence or above then n/a.
- Tracking, classify would help cue but wouldn't replace the analyst.
- PVO detector: useful because especially real-time, tactical commander level for decisions
- PVO tracker: useful for some level

- PVO detector: now distracting. Helpful if brought attention to other area of screen but couldn't read quickly.

- SOPs
- Normally they watch first, dictate (verbiage), and then make slide show. [E.g. (take screenshot… write description)*repeat]
- Traditionally they lay the image first, overlay the template next, and then save all as one file.
- [In AVAA], added videos to playlist (then did not have to scroll for screenshot renames).

List of Symbols, Abbreviations, and Acronyms

AEP	auditory-evoked potential
ANOVA	analysis of variance
AVAA	Advanced Video Activity Analytics
CV	computer vision
D	demand
DC	demand rating: complexity
DCGS-A	Distributed Common Ground Station-Army
DI	demand rating: instability
DV	demand rating: variability
ECG	electrocardiography
EEG	electroencephalography
ERP	event-related potential
FMV	full motion video
IR	infrared
ISIS	Islamic State in Iraq and Syria
KLV	key-length-value
MOS	military occupational specialty
NASA	National Aeronautics and Space Administration
PVO	person-vehicle-object
S	supply
SAR	supply rating: arousal-readiness
SART	Situation Awareness Rating Technique
SC	supply rating: concentration
SD	standard deviation
SS	supply rating: spare mental capacity

TLX	Task Load Index
U	understanding
UF	understanding rating: familiarity
UIQ1	understanding rating: information quantity
UIQ2	understanding rating: information quality
USB	universal serial bus
VBIED	vehicle borne improvised explosive device
V-NIIRS	Video National Imagery Interpretation Rating Scale

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIR ARL IMAL HRA RECORDS MGMT RDRL DCL TECH LIB
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
2 (PDF)	ARL RDRL HRB DG K M SCHWEITZER RDRL HRF C A J RIES

INTENTIONALLY LEFT BLANK.